Übungen zur Einf. in Deep Learning für Visual Computing

Sommersemester 2023 **Übungsgruppe 3**
Dozent: Prof. Dr. Reinhard Klein Ilia Mirkis
Tutorin: Alina Pollehn

Übungsblatt 1 **26.04.2023**

Aufgabe a)

Case: Regression von Surface Reflectance Parametern

Klasse von Aufgaben T

Eine Klasse von Aufgaben bezüglich dieser Problemstellung besteht aus Computationsaufgaben, die Images als Eingabe bekommen und verschiedene numerische Reflectance-Parametern (z.B. Reflexionskoeffizient), die dem Image entsprechen, calculieren können.

Performanz-Maß P

Zur Regression können wir MSE benutzen:

$$\mathrm{MSE}(\hat{\theta}_m) = \mathbb{E}\left[(\hat{\theta}_m - \theta)^2\right]$$

Erfahrung E

Wir können *supervised learning algorithm* benutzen. Als Dataset wird die Menge der Paaren ("Element", "Erwartungswert") benutzt.

Case: Klassifikation von MNIST Bildern

Klasse von Aufgaben T

Eine Klasse von Aufgaben bezüglich dieser Problemstellung besteht aus Entscheidungsaufgaben, die entscheiden, zu welcher von k Kategorien eine Eingabe gehört.

Performanz-Maß P

Zu den Klassifikationsproblemen können wir logloss Funktion (äq. Kreuz-Entropie) benutzen (als Ergebnis wird NN einen Wahrscheinlichkeitsvektor ausgeben):

$$\text{Logloss} = -\frac{1}{N}\sum_{i=1}^{N} y_i \ln(p_i) + (1-y_i)\ln(1-p_i)$$

Erfahrung E

Wie im obigen Case kann man als Dataset die Menge der Paaren ("Element", "richtige Kategorie") benutzen, um unser NN zu trainieren.

Aufgabe b)

Teil 1

$$P(X=i, Y=j) = \begin{cases} p(i)^2, \text{ falls } i=j \\ 2 \cdot p(i) \cdot p(j), \text{ falls } i<j \\ 0, \text{ falls } i>j \text{ (weil } \min \leqslant \max) \end{cases}$$

$$P_X(i) = p(i)^2 + 2p(i) \cdot \sum_{j=i+1}^{6} p(j) \implies \begin{cases} P_X(1) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot (4 \cdot \frac{1}{10} + \frac{1}{2}) = \frac{19}{100} \\ P_X(2) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot (3 \cdot \frac{1}{10} + \frac{1}{2}) = \frac{17}{100} \\ P_X(3) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot (2 \cdot \frac{1}{10} + \frac{1}{2}) = \frac{15}{100} \\ P_X(4) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot (1 \cdot \frac{1}{10} + \frac{1}{2}) = \frac{13}{100} \\ P_X(5) = \left(\frac{1}{10}\right)^2 + 2 \cdot \frac{1}{10} \cdot \frac{1}{2} = \frac{11}{100} \\ P_X(6) = \left(\frac{1}{2}\right)^2 = \frac{1}{4} \end{cases}$$

$$P_Y(i) = p(i)^2 + 2p(i) \cdot \sum_{j=1}^{i-1} p(j) \implies \begin{cases} P_Y(1) = \left(\frac{1}{10}\right)^2 = \frac{1}{100} \\ P_Y(2) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot \frac{1}{10} = \frac{3}{100} \\ P_Y(3) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot (2 \cdot \frac{1}{10}) = \frac{5}{100} \\ P_Y(4) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot (3 \cdot \frac{1}{10}) = \frac{7}{100} \\ P_Y(5) = \left(\frac{1}{10}\right)^2 + 2 * \frac{1}{10} \cdot (4 \cdot \frac{1}{10}) = \frac{9}{100} \\ P_Y(6) = \left(\frac{1}{2}\right)^2 + 2 \cdot \frac{1}{2} \cdot (5 \cdot \frac{1}{10}) = \frac{3}{4} \end{cases}$$

$$P(X=6 \mid Y=6) = \frac{P(X=6, Y=6)}{P(Y=6)} = \frac{\left(\frac{1}{2}\right)^2}{\frac{3}{4}} = \frac{1}{3}$$

$$P(X=1 \mid Y=3) = \frac{P(X=1, Y=3)}{P(Y=3)} = \frac{2 \cdot \frac{1}{10} \cdot \frac{1}{10}}{\frac{5}{100}} = \frac{2}{5}$$

$$P(Y=3 \mid X=1) = \frac{P(X=1 \mid Y=3) \cdot P(Y=3)}{P(X=1)} = \frac{\frac{2}{5} \cdot \frac{5}{100}}{\frac{19}{100}} = \frac{2}{19}$$

$$P(X=6 \mid Y=2) = \frac{P(X=6, Y=2)}{P(Y=2)} = \frac{0}{\frac{3}{100}} = 0$$

Teil 2

$$P(A) = \frac{N(A)}{N_{\text{total}}} \implies P(X=6 \mid Y=6) = \frac{N(X=6, Y=6)}{N(Y=6)} = \frac{4}{12} = \frac{1}{3}$$

Analog können wir andere Wahrscheinlichkeiten finden:

$$P(X=1 \mid Y=3) = \frac{N(X=1, Y=3)}{N(Y=3)} = \frac{2}{5}$$

$$P(Y=3 \mid X=1) = \frac{N(X=1, Y=3)}{N(X=1)} = \frac{2}{19}$$

$$P(X=6 \mid Y=2) = \frac{N(X=6, Y=2)}{N(Y=2)} = \frac{0}{5} = 0$$

Aufgabe c)

Zuerst beweisen wir, dass $\text{MSE}(\hat{\theta}_m) = \text{Bias}(\hat{\theta}_m)^2 + Var(\hat{\theta}_m)$

$$\text{MSE}(\hat{\theta}_m) = \mathbb{E}\left[(\hat{\theta}_m - \theta)^2\right] = \mathbb{E}\left[\hat{\theta}_m^2 - 2\hat{\theta}_m \cdot \theta + \theta^2\right] = \mathbb{E}\left[\hat{\theta}_m^2\right] - 2\theta \cdot \mathbb{E}\left[\hat{\theta}_m\right] + \theta^2$$

$$\text{Bias}(\hat{\theta}_m)^2 + Var(\hat{\theta}_m) = \left(\mathbb{E}\left[\hat{\theta}_m\right] - \theta\right)^2 + \mathbb{E}\left[\hat{\theta}_m^2\right] - \mathbb{E}\left[\hat{\theta}_m\right]^2 = \mathbb{E}\left[\hat{\theta}_m\right]^2 - 2\theta \cdot \mathbb{E}\left[\hat{\theta}_m\right] + \theta^2 + \mathbb{E}\left[\hat{\theta}_m^2\right] - \mathbb{E}\left[\hat{\theta}_m\right]^2 =$$
$$= \mathbb{E}\left[\hat{\theta}_m^2\right] - 2\theta \cdot \mathbb{E}\left[\hat{\theta}_m\right] + \theta^2$$

Da die rechten Seiten gleich sind, folgt daraus die ursprüngliche Behauptung. □

Diese zwei Gründen sind *Underfitting* und *Overfitting*.

Underfitting kann auftreten, falls unser Modell zu geringe Kapazität hat (sehr einfach ist). In diesem Fall ist der Bias hoch, weil die wahren Parametern θ nicht korrekt durch $\hat{\theta}$ approximiert werden können.

Overfitting kann auftreten, falls unser Modell zu hohe Kapazität hat ("too complicated" ist). In diesem Fall wird die Varianz hoch, weil es viele verschiedene Möglichkeiten für $\hat{\theta}$ gibt, die selben Daten zu approximieren.

Modelkapazität häangt von Hyperparameter ab. Deshalb ist es beim Wahl von Hyperparamatern sehr wichtig, Bias/Varianz Trade-off zu finden, weil es zur bessere Generalizierung führt.